J Cult Cogn Sci (2024) 8:309–333
https://doi.org/10.1007/s41809-024-00155-5

RESEARCH PAPER

Prediction ability in bilingual individuals: an eye tracking study with younger and older adults

Ingeborg Sophie Ribu · Hanne Gram Simonsen · Monica Norvik · Minna Lehtonen · Jeanett Murstad · Ane Theimann · Thomas Nygreen · Mira Goral

Received: 26 June 2024 / Revised: 25 October 2024 / Accepted: 29 October 2024 / Published online: 18 November 2024

Abstract Studies with monolingual speakers show that people predict upcoming linguistic elements during sentence processing. Linguistic prediction behavior has been found to be less consistent in studies with bilingual individuals performing in their non-native language and in neurotypical older monolingual adults. The present study utilized an eye-tracking paradigm to investigate whether bilingual younger and older neurotypical individuals predict upcoming nouns in sentences that include constraining verbs, and if they do so both in their first language (L1) and in their second language (L2). Data were analyzed from 44 Norwegian-English proficient bilingual adults; 27 younger (20–35 years, mean age 27) and 17 older adults (54–81 years, mean age 64) who completed the eye-tracking experiment in each of the two languages, as well as cognitive and linguistic tests. The results demonstrated similar prediction abilities in L1 and L2 for both the younger and older participants on sentences with constraining verbs. Older adults predicted slower than younger adults. Participants' working memory span and language proficiency did not explain prediction performance; cognate status of the stimuli partially did. The study adds to the relatively sparse existing data on prediction abilities in bilingual people and in older individuals.

Keywords Eye-tracking · Bilingual · Constraining verbs · Older adults · Norwegian

I. S. Ribu (✉)
Department of Vocational Teacher Education, Oslo Metropolitan University, Oslo, Norway
e-mail: ininge@oslomet.no

H. G. Simonsen · A. Theimann
Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway

M. Norvik
Department of Education, UiT The Arctic University of Norway, Tromsø, Norway

M. Lehtonen
Department of Psychology and Speech Language Pathology, University of Turku, Turku, Finland

J. Murstad
Stockholm University, Stockholm, Sweden

T. Nygreen
The Norwegian Railroad Directory, Oslo, Norway

M. Goral
Speech-Language-Hearing Sciences, Lehman College & Graduate Center, The City University of New York, New York, NY, USA

Introduction

Linguistic prediction can be defined as pre-activation of upcoming linguistic information based on certain cues in the language (e.g., Huettig, 2015). Studies with monolingual speakers show that individuals predict upcoming linguistic elements during sentence processing (e.g., Altmann & Kamide, 1999; Kuperberg & Jaeger, 2016; Yoshida et al., 2013). In a seminal eye-tracking study, Altmann and Kamide (1999) discovered that monolingual individuals moved their eye gaze towards specific objects (e.g., 'cake') before they were mentioned when listening to constraining verbs (e.g., 'eat'), but not when listening to neutral verbs (e.g., 'move'). Linguistic prediction behavior has been found to be less consistent in studies with bilingual individuals performing in their non-native language, with studies showing that bilingual people have limited ability to predict in their second language (e.g., Ito et al., 2018). Moreover, mixed results have also been reported for neurotypical older monolingual adults (e.g., Huettig & Janse, 2016; Maquate & Knoeferle, 2021). In the current study we utilized an eye-tracking paradigm to investigate whether bilingual, neurotypical younger and older individuals predict upcoming nouns in sentences that include constraining verbs (e.g., an edible object such as 'pizza' after a verb such as 'eat') in a similar manner as monolingual individuals do, and if they do so both in their first language (L1) and in their second language (L2).

Prediction studies with bilingual people

Researchers have asked whether prediction abilities in the two languages of bilingual individuals—those who use more than one language in their lives—are affected by variables such as age of second language acquisition, language proficiency, and language-specific characteristics. Grüter et al. (2014) proposed that when people listen in their non-native language, their ability to generate expectations is lower than when listening in their native language. The Reduced Ability to Generate Expectations (RAGE) hypothesis (Grüter et al., 2014) suggests that in their non-native language, bilingual people will predict upcoming information at a slower speed compared to their native language or compared to monolingual people. Consistent with the RAGE hypothesis, a recent review article (Schlenter, 2023) concluded that prediction studies comparing L1 and L2 processing almost always find later prediction onset times and weaker effects for processing in L2 compared to L1.

Prediction studies have documented differences between predictive processing in L1 and L2 for a variety of language pairs and a range of predictive cues. These differences include not using certain cues to predict upcoming information or using them less efficiently. For instance, English-speaking L2 learners of Spanish did not predict upcoming nouns based on gender-marked articles, while L1 speakers of Spanish did (Lew-Williams & Fernald, 2010). In another study, experienced L2 learners of Spanish did predict nouns based on the gender of the articles like L1 speakers, but their prediction was slower (Grüter et al., 2012). Additional studies showed that Chinese learners of English processed prosodic cues (contrastive pitch accent) but did not then use them to predict upcoming referents in their L2 (Perdomo & Kaan, 2021), whereas L1 speakers of English did. Similarly, English speaking L2 learners of German did not predict upcoming information based on case markings in German, while L1 German speakers did. In turn, the L2 learners predicted upcoming information based on semantic cues, indicating that linguistic prediction may be different for L1 and L2 speakers (Hopp, 2015), especially for people who are not highly proficient in their L2. Schlenter (2023) brings up the possibility that differential cues across languages (e.g., whether nouns have gender marking or not, or whether thematic roles are marked with case or word order) can explain different prediction patterns, in particular for grammar-based cues. This might suggest a role of L2 proficiency and exposure in prediction performance, as L2 learners with a higher proficiency (or more L2 exposure) may internalize and use relevant L2 cues instead of overusing other, often L1-based cues (e.g., word order and/or semantics instead of morphological case marking).

Indeed, studies found that proficient bilingual speakers' prediction abilities were similar to those of their monolingual counterparts. For example, Dutch–English proficient bilingual participants predicted upcoming nouns based on constraining verbs in both their languages, but were slower in both languages compared to English monolingual participants (Dijkgraaf et al., 2017). Moreover, another group of Dutch–English proficient bilinguals predicted semantically related nouns when hearing constraining verbs in both languages, but the effect size was larger in Dutch (L1) than in English (L2) (Dijkgraaf et al., 2019). English L2 speakers of various L1 backgrounds performed similarly to English L1 speakers in an eye-tracking experiment using orally presented sentences with constraining verbs, but were slower when cognitive load was manipulated (Ito et al., 2018, see below).

Kim and Grüter (2021) found that Korean learners of English were slower at predicting in their L2, whereas Contemori and Dussias (2019), who studied early Spanish–English bilinguals, found no difference in speed of prediction compared to monolingual English speakers. They argued that the participants in their study were able to predict referents similarly to monolingual English speakers, since the bilingual participants in their sample were early bilinguals (Contemori & Dussias, 2019).

Higher language proficiency and earlier age of acquisition may entail more stable lexical representation in that language. Prediction ability has been associated with lexical knowledge (e.g., Borovsky et al., 2012; Federmeier et al., 2002; Rommers et al., 2015). That is, a stable vocabulary representation may be needed for listeners to be able to predict, which may explain why less proficient L2 users do not predict upcoming information (e.g., Pickering & Gambi, 2018). For example, Rommers et al. (2015) found that high receptive vocabulary and high category fluency scores were associated with higher prediction ability in their Dutch speaking participants. In contrast, Corps et al. (2023) concluded from their eye-tracking experiment with advanced English learners (of various L1 backgrounds) that L2 proficiency did not mediate prediction, however, their L2 participants were all highly proficient.

A potential way to explore the effect of strength of lexical representation in bilingual language processing may be to compare words that are cognates (sharing meaning and form) and non-cognate translation equivalents. It is well established through bilingual word recognition studies that cognates make lexical access easier—cognates are recognized or produced faster than control words in lexical decision or naming tasks (e.g., Costa et al., 2000; Dijkstra et al., 2010; Jared & Kroll, 2001; Martinez-Garcia, 2019). Such a cognate facilitation effect is ascribed to coactivation of the cognate words during processing (non-selective access), and stronger lexical representations also due to higher subjective frequency since cognates are heard and used in both languages (Blumenfeld et al., 2016). However, few prediction studies thus far examined this aspect, and findings are mixed (Van Assche et al., 2012). Schwartz and Kroll (2006) tested cognate effects in Spanish–English bilinguals in a sentence prediction reading task where the target nouns had to be named, and found cognate facilitation in low-constraint sentences (sentence frames that do not bias the target word), but not in high-constraint sentences (sentence frames that bias the target word). Duyck et al. (2007) used eye-tracking to investigate cognate facilitation in Dutch–English bilinguals, in low-constraint sentences. They found significant cognate facilitation across all eye-tracking measures, but only for identical cognates—the size of the cognate effect increased as a function of lexical similarity between targets and translation equivalents. Most of such investigations involve nouns only; one exception is Van Assche et al. (2013), who found a smaller and later cognate effect in sentence processing using verb stimuli than generally found with noun stimuli.

Other, non-linguistic variables that have been associated with prediction ability include speed of processing and working memory (e.g., Huettig & Janse, 2016; Ito et al., 2018; Li & Qu, 2024). Huettig and Janse (2016) collected eye-tracking data from 105 L1 speakers of Dutch between the ages of 32 and 77. In their regression analyses they demonstrated that when controlling for age, working memory (WM), as measured by non-word repetition, digit span backwards, and Corsi block tasks, as well as processing speed, as measured by digit-symbol substitution and letter comparison tasks, predicted anticipatory eye movements based on article gender cues. Ito et al. (2018) examined eye-tracking patterns of L1 and L2 users when they listened to sentences and viewed target and non-target images compared to when they performed a memory task in addition to the listening task. The authors found that whereas the participants who performed the task in their L2 had similar prediction behavior to the L1 participants in the listening only condition, their eye movements were delayed in the cognitively taxing condition (Ito et al., 2018). Similarly, Li and Qu (2024) found that Chinese speaking participants with higher verbal WM spans showed earlier prediction times than participants with lower WM spans. In addition, age, which correlated with both speed and WM, emerged as a significant predictor, as will be discussed next.

Prediction studies with older adults

Prediction ability has not been studied extensively in older age and results to date are mixed (Maquate & Knoeferle, 2021). Older age has been associated with compromised cognitive abilities such as processing speed and WM (e.g., Salthouse, 1991), the same abilities that have been found to modulate prediction behavior. Evidence for slower prediction in older age can be found in the study by Maquate and Knoeferle (2021), who showed that the older adults in their study had similar fixation patterns as the younger adults; that is, they fixated more on the agent than the competitor when the action was depicted. However, the fixations started later for the older than for the younger adults.

However, research has demonstrated that older adults may use top-down processing more than younger adults do, perhaps to compensate for reduced sensory processing, and thus demonstrate better prediction ability (Federmeier et al., 2010). Indeed, Huettig and Janse (2016) demonstrated that older age, when dissociated from cognitive abilities, was associated with more anticipatory eye movements. In another study, a self-paced reading paradigm with younger and older monolingual adults revealed that older age correlated with faster responses related to bigrams with higher transition probability (i.e., the probability of occurrence in a given context), implying, that older adults with more experience and larger vocabularies have more entrenched probabilistic knowledge (McConnell & Blumenthal-Dramé, 2021).

The association of better lexical abilities and older age fits the approach promoted by Ramscar, Baayen and colleagues (Ramscar & Baayen, 2015; Ramscar et al., 2014). They suggest that given that learning continues throughout the lifespan, the increasing amount of information that needs to be processed, for example as a result of the increase in vocabulary size, needs to be considered when accounting for observed slower or less efficient processing. For example, a greater number of competitors during lexical selection may be activated for individuals with larger rather than smaller vocabularies.

The current study

As the brief review above suggests, mixed results have been reported for the prediction behavior of older adults and of bilingual adults. In L2 users, language proficiency emerged as a critical variable for prediction ability. It is likely that some of the mixed results reported for prediction behavior in the L2 may be explained by considering the L2 proficiency of the participants enrolled. Moreover, few studies compared prediction performance in the two languages of bilingual individuals directly, rather, in most cases, processing in the L2 of bilingual groups has been compared to processing of monolingual speakers of that language. Furthermore, the role of lexical similarity between the two languages has not been addressed much in the literature, and none of these studies have examined prediction performance in older bilingual adults.

With this in mind, the aim of the present study was therefore to examine the prediction ability of younger and older proficient sequential bilingual individuals during processing in their L1 and in their L2, by means of the visual world paradigm, which is particularly well suited to study online language processing (Huettig, 2015). We asked the following research questions:

1) Do people predict upcoming language (specifically, nouns following constraining and neutral, non-constraining verbs) as they hear sentences in their L1 and in their L2?
2) How does aging affect the prediction processing in L1 and in L2?
3) What is the influence of participants' language proficiency and cognitive abilities on the prediction process?
4) What is the influence of cognate status of the stimuli on the prediction process?

On the basis of previous research, we expected that bilingual people who were highly proficient in both languages would predict nouns following constraining verbs in both their L1 and L2, and that prediction abilities would be evident in younger and older participants. Larger vocabulary and richer language experience, associated with high proficiency and with older age, can contribute to better prediction ability. In contrast, reduced processing speed and memory capacity—also associated with older age and with processing in a non-native language—may contribute to reduced prediction ability. By including both proficiency and cognitive measures, we aimed to disentangle reasons for possibly differential prediction abilities of older vs. younger adults. In general, we expected L2 proficiency to be associated with better prediction performance, and that better cognitive abilities, such as working memory, would be associated with better prediction abilities. Finally, we predicted that cognates would be easier to predict than non-cognates, as they typically have stronger lexical representations due to shared form and meaning across languages and hence contribute to easier and faster lexical access.

Methods

Participants

Fifty-two participants were enrolled in the study. The eye-tracking experiment was part of a larger study on bilingual language processing. The full project also included an EEG study that is not reported here. After the experiment, eight participants were excluded due to poor eye tracking quality (six older and two younger participants). More data had to be discarded from the older than the younger participants, which may be related to physiological differences connected with aging (i.e. droopy eyelids and false corneal reflections that are more common in older adults (Holmqvist et al., 2011)), glare from glasses or technical issues with the eye tracking equipment. In the end, 44 participants were included in the study: 27 younger (age ranging between 20 and 35) and 17 older (ranging between 54 and 81 years of age) Norwegian–English bilingual individuals.

Participants were recruited through personal networks, flyers, posts on social media and the university website. Flyers were put up at the University of Oslo campus and in relevant places in and around Oslo (community centers, libraries, and bulletin boards). Participants were also asked to help recruit from their personal network after attending the experiment. Several potential participants that showed initial interest in the project declined to take part in the experiment due to time constraints. Originally, we aimed to include a minimum of 30 participants in each age group, based on sample sizes from similar studies, but due to difficulties recruiting a large enough sample we were left with 27 younger and 17 older participants after exclusion. An effect size sensitivity analysis and a power determination analysis (Giner-Sorola et al., 2024) was conducted to detect the smallest possible effect size and power given our final sample (see Sect. "Influence of age, education, language proficiency and working memory on predictive ability" below).

All participants had normal, or corrected to normal, hearing and vision. All participants spoke Norwegian as their first language and had a high command of English (many Norwegians have high proficiency in English and are exposed to English in their daily life). English proficiency was measured via verbal fluency tasks (category and letter fluency), an English-to-Norwegian word translation task, a word familiarity rating and self-rated proficiency for speaking, comprehension and reading (see Materials below). Age, education, language background information, and results on the cognitive and linguistic tasks are summarized in Table 1.

The participants gave written consent to participate in the study, and were compensated with a gift card. The study was approved by the Norwegian Agency for Shared Services in Education and Research (SIKT) before data collection started.

As can be seen from Table 1, there were differences between younger and older participants in regard to age of L2 acquisition, with younger adults having learned English earlier than the older participants, and self-rated proficiency in English; the younger participants rated themselves as more proficient than the older participants. The younger participants also had a smaller difference in verbal fluency scores between Norwegian and English. Regarding the cognitive tasks, there were differences between younger and older adults in reaction time on the Flanker task for congruent trials, incongruent trials and for the global reaction time (i.e., the mean reaction time from the congruent and incongruent trials). In all cases, the younger adults were faster than the older adults. Finally, there were differences between the older and younger participants on the Corsi forward and backward span tasks, where the younger participants remembered longer span sequences than the older participants. There were no differences in years of education, translation scores, word familiarity rating,

Table 1 Participant characteristics

	Younger (n = 27) Mean (SD)	Older (n = 17) Mean (SD)
Gender	19 F: 8 M	11 F: 6 M
Age	26.81 (4.53)	64.41 (8.86)
Range	20–35	54–81
Education (years)	17.11 (2.48)	15.05 (6.72)
Range	13–24	10–28
L2 age of acquisition	6.40 (2.85)	10 (2.87)
Range	0–17	0–13
Self-rated proficiency (composite score, max 10)	8.32 (1.08)	7.21 (2.21)
Range	5–10	5–10
Translation (max 250)	170.74 (113.05)	191.94 (92.03)
Range	138–250	13–246
Familiarity rating (max 250)	234.7 (58.22)	210.82 (90.36)
Range	72–252	72–252
Verbal fluency English	43.22 (9.55)	39.02 (9.87)
Range	26–65	21–63
Verbal fluency Norwegian	51.64 (11.30)	53.7 (11.46)
Range	34–76	35–82
Verbal fluency difference Norwegian–English	8.42 (9.07)	14.67 (6.67)
Range	−19–27	6–28
Digit span forward	5.55 (2.29)	6.23 (0.91)
Range	0–8	5–8
Digit span backwards	4.51 (2.22)	4.94 (1.59)
Range	0–8	3–8
Flanker congruent reaction time	383.03 ms (50.69)	481.35 ms (99.37)
Range	314–522 ms	384–727 ms
Flanker incongruent reaction time	462.05 ms (61.52)	574.63 ms (132.83)
Range	383–608 ms	441–880 ms
Flanker global reaction time	422.54 ms (54.95)	527.99 (112.84)
Range	348–562 ms	413–748 ms
Flanker effect	79.01 ms (25.10)	93.28 ms (64.11)
Range	41–134 ms	23–287 ms
Flanker SCE	2.23 (24)	2.60 (39.33)
Range	−49–34	−71–76
Corsi forward span	6.74 (1.05)	5.05 (0.96)
Range	5–9	3–6
Corsi backwards span	6 (1.44)	4.76 (1.34)
Range	4–9	3–8

Given that inferential statistics is meant to infer something about a specific population, and not a sample of that population, the differences between younger and older adults in the current sample are presented descriptively rather than through testing for significance between groups (Sassenagen & Alday, 2016)

verbal fluency in English, verbal fluency in Norwegian, or digit span tasks (forwards and backwards).

Procedure

Participants were tested individually in two sessions of about 1.5 h each. The sessions were administered on two separate days; one day when all tasks were completed in Norwegian and the other day when all tasks were completed in English. Each participant completed a background questionnaire, the eye-tracking experiment (see Sect. "Eye-tracking experiment") in Norwegian and English, verbal fluency tasks in each language and three cognitive tasks (digit span tasks, Flanker and Corsi block tasks) (see Sect. "Linguistic and cognitive tasks"). At the end of the last session, participants completed a word familiarity task and a translation task. Sessions were audio- and video-recorded. The order of the cognitive and linguistic tasks and language of administration were counterbalanced across participants. On the days with the Norwegian eye-tracking experiment, participants performed the verbal fluency task in Norwegian, the digit span tasks and either the Flanker or the Corsi block task. On the day of the English eye-tracking task, the participants completed the verbal fluency task in English and either the Corsi block task or the Flanker task depending on which they did not complete during the Norwegian test session.

Tasks and materials

Eye-tracking experiment

Participants completed a visual world paradigm experiment in which they were instructed to listen to sentences and look at pictures on a computer screen. Apart from this, there was no explicit task they needed to perform. The stimuli were presented in a random order, yielding a different order of presentation for each participant. Each participant listened to a total of 72 experimental sentences after a calibration and validation trial. They were given two short breaks after approximately each third of the trials.

The auditory stimuli consisted of 36 sentences with constraining verbs and 36 sentences with neutral verbs, all in the present tense, each presented after an introductory sentence (e.g., *This is Anna. Anna eats Mary's pizza)* in each language. (See Table 2 for examples of the stimuli.). Verbs in any English-Norwegian verb pair were of comparable frequency and imageability. Frequency and imageability ratings for Norwegian were taken from the database *Norwegian Words* (Lind et al., 2015), English frequency data came from the SUBTLEXus database (Brysbaert et al., 2012) and imageability ratings in English were obtained from the Glasgow norms (Scott et al., 2019). Additionally, half of the verbs in each condition were cognates (e.g., *plants* in English; *planter* in Norwegian) and half of the target nouns in each condition were cognates too (e.g., *door* in English; *dør* in Norwegian), yielding four categories of 18 verbs (constraining-cognate; constraining-non-cognate; neutral-cognate; neutral-non-cognate), and four categories of 18 nouns (constraining-cognate; constraining-non-cognate; neutral-cognate; neutral-non-cognate); see Table 2.

Cognate status was established through "Researcher Intuition", following Strangmann et al. (2023). Strangmann et al. (2023) compared three different methods of cognateness evaluation: Levenshtein Distance, Translation Elicitation, and Researcher Intuitions, and found moderate to high correlation among them. We chose Researcher Intuition since this is the only method assessing both form and content overlap. Two of the authors (MN and HGS), proficient Norwegian-English bilinguals, separately rated the degree of overlap of the word pairs on a seven-point scale, 0 indicating no overlap and 6 indicating full overlap. The raters were in nearly full agreement, and in the very few cases where the scores differed (never by more than one point difference), consensus was reached through discussion. Word pairs were counted as non-cognates with a composite rating of 0–1, and as cognates with a rating of 3–6. We did not distinguish between identical and non-identical cognates in this study, since there are very few identical cognates between Norwegian and English due to orthographic and morphological differences between the languages. In our dataset, there are no identical verbs, and only four identical nouns (*pizza, rose, pipe, glass*), and among these two of them (*rose, pipe*) are only orthographically identical.

The visual stimuli consisted of a set of three pictures (one fixation picture, one target and one distractor picture), which were introduced 1000 ms before the onset of the spoken sentence. All pictures were black and white line drawings presented on a white background. Some of the drawings were created by Marko Belić for the Comprehensive Aphasia Test (Swinburn et al., 2021), and others were from the

Table 2 Example sentences

	Norwegian	English
Constraining-cognate verb (n = 18)	Anne PLANTER Maris tulipan	Mary PLANTS Anna's tulip
Constraining-non-cognate verb (n = 18)	Anne MATER Maris papegøye	Mary FEEDS Anna's parrot
Neutral-cognate verb (n = 18)	Anne LIKER Maris rose	Anna LIKES Mary's rose
Neutral-non-cognate verb (n = 18)	Anne VELGER Maris genser	Mary CHOOSES Anna's sweater
Constraining-cognate noun (n = 18)	Anne drikker Maris KAFFE	Mary drinks Anna's COFFEE
Constraining-non-cognate noun (n = 18)	Mari strikker Annes LUE	Anna knits Mary's HAT
Neutral-cognate noun (n = 18)	Anne finner Maris TE	Mary finds Anna's TEA
Neutral-non-cognate noun (n = 18)	Mari møter Annes HEST	Anna meets Mary's HORSE

MultiPic database (Duñabeitia et al., 2018) and the Clipart Library (www.clipart-library.com), with a free for commercial use creative commons license. Pictures of the 36 nouns were used in all conditions such that each participant saw each picture twice as a target (once following a constraining verb and once following a neutral verb) and twice as a distractor across the complete set of 72 trials, in each language. A picture of the subject of the sentence, the fixation picture, (Anna/Anne or Mary/Mari) was placed on the top middle of the screen, and the target (e.g., 'pizza') and distractor (e.g., 'cup') were placed under the fixation picture to the left or right side of the screen. Half of the sentences included a picture of 'Anna/Anne', and half a picture of 'Mary/Mari' (see Fig. 1). The placement of the target picture to the left or right was counterbalanced across trials.

The stimuli were presented on a 22'' external display (1920 × 1080 pixels). Participants were seated between 50 and 70 cm from the screen. The fixation images measured 800 × 800 pixels, and the distractor and target images measured 1000 × 1000 pixels. These pictures also corresponded to the three main areas of interest (AoI); fixation picture, target and distractor. A fourth AoI, white space, was assigned to the rest of the screen where no visual stimuli was present.

The auditory stimuli were recorded by a Norwegian–English early simultaneous bilingual female speaker with a Zoom Q2n recorder, with 48,000 Hz. The stimuli were edited in Audacity version 2.2.2 (Audacity Team, 2018). A short pause (mean 812 ms, *SD* 210 ms) was added after the context sentence, to ensure that the verb onset was at exactly 3500 ms for all sentences. Another short pause (mean 564 ms, *SD* 143) was included after the verb, so the noun onset was at 5300 ms (e.g., *This is Anna*. [pause] *Anna eats* [pause] *Mary's pizza*). The predictive window is set between the onset of the verb and the onset of the noun plus a 250 ms buffer, which corresponds to the time it takes to launch a saccade and land a fixation as a response to a stimulus, making the length of the predictive window 2050 ms long. The buffer was set at 250 ms, since the latency of saccadic movements increases with age (Pelak, 2010). No linguistic cues appeared during the predictive window that could bias the participants to look towards either of the pictures.

Auditory stimuli were presented through external speakers, connected to the monitor. The experiment was controlled and monitored from a laptop computer. The procedure started with a five-point calibration and validation of both eyes. If there was a deviance of more than 0.5 degrees between calibration and validation, a second calibration was conducted. Each trial was preceded by a fixation cross in the middle of the screen. The fixation cross was left on the screen for 1000 ms before the experimental stimuli appeared. The experimenter would manually move on to the next trial after approximately 3 s.

Participants' eye gaze from both eyes was measured with an SMI eye-tracker RED250MOBILE, a remote eye-tracking device with a sampling rate of 250 Hz, i.e., approximately every 4 ms. We excluded

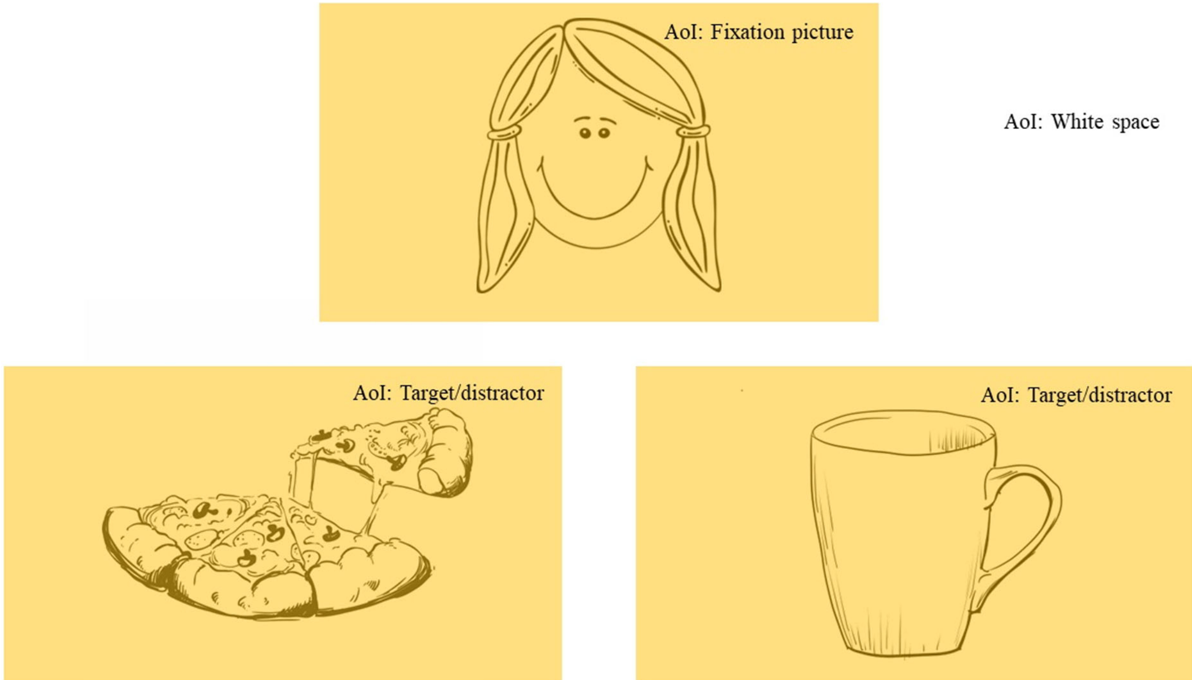

Fig. 1 The visual display, with areas of interest

trials if participants indicated on the word familiarity task that they did not know the meaning of the verb or the noun in the sentence, and all trials where the tracking ratio fell below 75% due to blinks, track losses or gaze drifts off the screen. After cleaning the data, an average number of approximately 35 analyzable trials were left per condition, per language and per age group (Table 3).

Table 3 Number of analyzable sentences per condition and age

	Constraining		Non-constraining	
	Norwegian	English	Norwegian	English
Younger	35.14	35.7	36.3	34.4
Older	35.1	34.4	35.8	34.2

Linguistic and cognitive tasks

The participants filled out a language background questionnaire (language experience and proficiency questionnaire (LEAP-Q) (Marian et al., 2007). They also completed a five-part verbal fluency task in each language: two category fluency (Animals; Verbs) and three phonological fluency (FAS) (Spreen & Benton, 1969). In addition, participants completed three cognitive tasks: a Flanker task (adapted from Eriksen & Eriksen, 1974), a Corsi block task, forward and backward (see Waris et al., 2015), and a digit span task, forward and backward (Waris et al., 2015). Finally, after completing the experiments, they were administered two lexical proficiency tasks in English. First, a word familiarity rating, in which the participants indicated how well they knew the meaning of each English word included in the stimulus sentences on a four-point scale (*I am confident that I know the meaning of this word*, *I somewhat know the meaning of this word*, *I am uncertain about the meaning*, and *I don't know the meaning of this word*). Second, the participants performed a translation task, in which they translated the same words as in the familiarity task, in a random order, from English to Norwegian, using a link to a web-based translation task. The online translation task was hosted on the eBabyLab platform (Lo et al., 2023). The results from the online task were stored anonymously on a secure University of Oslo server.

Data analysis

A multilevel logistic regression (Barr, 2008) was fitted to investigate whether there was a difference in proportions of looks to the target image during the predictive window, in L1 and L2 and between conditions. Next, following Stone et al. (2021) we conducted a divergence point analysis using a bootstrapping approach for the eye-tracking data to pinpoint the time when there is a sustained effect of more looks to the target over the distractor (see section Analysis of eye-tracking data). Furthermore, to investigate the effects of age, language proficiency and cognition on predictive ability, correlation analyses and a multiple linear regression analysis were conducted (see section Analysis of cognitive and background variables).

Analysis of eye-tracking data

All analyses were performed in R 4.3. × 3 (R Core Team, 2024), using RStudio, version 2024.04 (Posit team, 2024). We used the following packages: *boot* (Canty & Ripley, 2024; Davison & Hinkley, 1997) *car* (Fox, 2019), *corrplot* (Wei & Simko, 2021), *data.table* (Barrett et al., 2024), *dtplyr* (Wickham et al., 2023a, 2023b), *Hmisc* (Harell, 2024), *lme4* (Bates et al., 2015), *progress* (Csárdi & FitzJohn, 2023), *psych* (Revelle, 2024), *R.utils* (Bengtsson, 2023), *rmarkdown* (Allaire et al., 2024), *scales* (Wickham et al., 2023a, 2023b), *pwr* (Champely et al., 2022), and *tidyverse* (Wickham et al., 2019).

Fixations to target were binary coded (1 = fixation, 0 = no fixation) for each AoI at each 4 ms sampling interval and placed in 20 ms time bins. Predictive ability was operationalized as the mean proportion of fixations to the target AoI divided by the total number of fixations to all AoIs within the predictive window (i.e., predictive looks) per participant (Mani & Huettig, 2012). The proportion of predictive looks to the target in the predictive window was used as the dependent variable in the further analyses in the regression models and the divergence point analysis. Similarly, the proportion of looks to the Distractor image was calculated taking fixations to the distractor AoI and dividing these by the total number of fixations to all AoIs within the predictive window. Proportions of fixations convey the proportion of time the eye gaze is stable in a given area on the screen, and is therefore commonly employed in visual world paradigms (Ito & Knoeferle, 2023).

In the bootstrapping function, the fixations to the White Space and Fixation Picture AoIs were removed, and only fixations to the target or the distractor were included for analysis. The divergence point, i.e., the time point when participants fixated significantly more to the target than to the distractor image, was identified as the first series of 10 consecutive 20 ms time bins (for a total of 200 ms) with a *p value* less than 0.05 on a one-sided t-test. The test used the proportion of fixations to the target image among the fixations to either the target or distractor images, with a null hypothesis that the proportion was less than or equal to 0.5. While Stone et al. (2021) test samples containing the mean fixation proportion per participant within the time bin, we test samples containing the mean fixation proportion per trial within the time bin. This avoids an unnecessary level of aggregation and increases the statistical power. As individual observations within the same time bin in the same trial are likely to be to the same region, the sampled proportions will mostly be 0 or 1. A binomial test would model these well, but would not handle the minority of intermediate proportions. As a consequence, they would have to be rounded up or down, losing the differences between a slight and full majority of looks to one region. However, around the divergence point, the sample *mean*, reflecting the probability of a fixation to the target, will still be distributed around 0.5, where a t-test approximates a binomial test well. A linear model with weighted logits (Barr, 2008; Vanek et al., 2024; Veríssimo & Clahsen, 2014) is not recommended for fitting models because it creates spurious interactions (Donnelly & Verkeulen, 2017).The results from the alternative binomial test and the logistic regression with weighted logits in the bootstrapping analysis, and corresponding time series plots, can be found in Supplementary materials on OSF: https://doi.org/10.17605/OSF.IO/VEMY4.

We then performed a bootstrapping analysis to estimate the distribution of the statistical test results if the experiment would be repeated several times; this means that an existing dataset is resampled to create 'new' datasets (Stone et al., 2021). We applied a non-parametric bootstrap using stratified resampling with replacements. To speed up the bootstrap, we used a customized function, based on the *boot* function in Canty & Ripley's *boot* package (2024). The data were resampled 2000 times in each analysis, maintaining the same number of observations per participant per time bin as in the original sample.

We performed three bootstrapped analyses; the first investigated the divergence point differences between younger and older participants in both L1 and L2; the second investigated the differences in divergence points between older and younger adults to sentences in L1 and L2 containing verbs that are cognates and non-cognates between the two languages, and finally, the third bootstrapping analysis investigated the difference in divergence points between older and younger adults on sentences where the target nouns were cognates and non-cognates between Norwegian and English.

Analysis of cognitive and background variables

The language proficiency measures consisted of a (1) composite score for the self-rated proficiency measures (speaking, comprehension and reading) from the LEAP-Q; (2) phonological verbal fluency (in each language); (3) category fluency (in each language); (4) a difference score between the verbal fluency composite score (phonological plus category) between L1 and L2; (5) the raw scores from the translation task (max. 250 points); and (6) the word familiarity rating (max. 250 points). Accuracy scores on the Corsi block task and digit span tasks (forward and backward) were included as measures of working memory, whereas the Flanker effect (difference between incongruent and congruent trials) was used to measure inhibition, and the Flanker sequential congruency effect (SCE) was included to measure conflict adaptation. The SCE score assumes that there is a smaller congruency effect between two incongruent trials than two congruent trials (Grundy, 2022). Mean reaction time on the congruent and incongruent trials in the Flanker tasks were recorded as a proxy for processing speed.

We first ran correlation analyses separately on the Norwegian and the English data to explore the relationship between the proportion of predictive looks to target and the cognitive and linguistic background measures. The subtasks from each cognitive and linguistic task that correlated best ($r > 0.3$ and $p < 0.05$) with the proportion of predictive looks to target during the predictive window on the sentences with constraining verbs (namely, the digit span backwards and the Flanker effect, and the category fluency for prediction in Norwegian; and the same plus verbal fluency difference, word familiarity and self-rated proficiency for prediction in English), were included in a multiple linear regression analysis to investigate the effect of cognitive and language functioning on predictive abilities in L1 and L2.

We fitted the first multiple linear regression model using the proportion of predictive looks as our response variable, and age, digit span backward, Flanker effect and category fluency scores in Norwegian as potential predictor variables for the Norwegian model.

For the second model, for prediction in English, the predictor variables in the model were age, digit span backward, Flanker effect, category fluency scores in English, as well as the difference in total verbal fluency scores between Norwegian and English, word familiarity and self-rated proficiency.

Results

Eye-tracking results

Difference in proportion of looks within the predictive window per condition and language

Predictive processing was analyzed by means of a multilevel logistic regression analysis (Barr, 2008; Garrido Rodriguez, 2023) with fixations to the target within the predictive window as the dependent variable, and condition (neutral vs. constraining verbs), language (L1/Norwegian vs. L2/English) and participant as random intercepts. There was a significant difference in the proportion of looks to target images between conditions (neutral vs. constraining) and language (Table 4), indicating more predictive looks to the target in the constraining condition than in the neutral condition.

Divergence point analysis for sentences with constraining verbs

In the following parts, the divergence points for the sentences with constraining verbs are reported. The divergence point indicates when there is a sustained effect of more looks to the target over the distractor. Since the sentences with neutral verbs do not contain any predictive information, it is not expected that the divergence point for these sentences will fall within the predictive window. The divergence point for the sentences with neutral verbs are only reported to show that there are no predictive looks in sentences with non-constraining verbs. All time values are milliseconds from the beginning of the predictive window (verb onset, not trial onset). The predictive window is therefore from 0 to 2050 ms.

Results from the bootstrapped divergence point analysis between younger and older participants are shown in Table 5 for all sentences with constraining and non-constraining verbs. For the sentences with constraining verbs, the younger participants fixated significantly more to the target than to the distractor at 1184 ms after the verb onset in L1 (with 95% confidence interval (CI) [1140, 1240 ms]), and 1188 (95% CI [1160, 1240 ms]) ms after the verb onset in L2. The divergence point for the older participants was 1495 ms after the verb onset for L1 (95% CI [1320, 1580 ms]) and 1635 ms after the verb onset for L2 (95% CI [1500, 1780 ms]). For both age groups the average divergence points are within the predictive window for sentences with constraining verbs. Fixation curves of the proportion of looks and the divergence points for the sentences with constraining verbs are presented in Fig. 2. For sentences with non-constraining verbs, all divergence points were outside of the predictive window, see Table 5.

The difference between younger and older participants on sentences with constraining verbs in the L1 was 312 ms (95% CI [140, 420 ms]) from the verb onset and 447 ms (95% CI [320, 600 ms]) from verb onset in L2, indicating a faster detection of the Target image for the younger participants in both languages. There was no difference in the time to fixate on the Target between L1 and L2 for the younger (5 ms; 95% CI [-60, 80]) nor the older (140 ms; 95% CI [-20, 340]) participants. Figure 3 shows the histogram and 95% CI for the difference in divergence points between trials in the L1 versus L2 for the older and younger participants.

For the sentences with neutral verbs, all participants showed preferential looking patterns; that is, more looks to the target relative to the distractor image, after the predictive window. This was true in both the L1 and the L2 condition. This was expected, since there are no linguistic cues that can aid prediction in these sentences, and these sentences served as a manipulation check to establish that the neutral verbs were in fact not providing prediction cues.

Table 4 Results from the multiple regression model for predictive looks between conditions and languages

	Estimate	sd	**z-value**	**p-value**
Intercept	−1.11	0.17	-6.31	< 0.001***
Language	−0.19	0.06	-2.87	0.004**
Condition	−0.87	0.06	-12.82	< 0.001***

**indicates a *p-value* between 0.05 – 0.001, and ***a *p-value* < 0.001

Divergence point analysis for sentences with cognate and non-cognate verbs and nouns

As shown in Table 6, the divergence point for sentences with cognate verbs collapsed across trials in both languages was at 1154 ms after the verb onset for the younger participants (95% CI [1120, 1200 ms]) and 1619 ms after the verb onset for the older participants (95% CI [1540, 1780 ms]). For sentences with non-cognate verbs, the divergence point for the younger participants was 1224 ms after the onset of the verb (95% CI [1180, 1260 ms]), and for the older 1462 ms after the onset of the verb (95% CI [1360, 1580 ms]). The fixation curves for sentences with cognate and non-cognate verbs are presented in Fig. 4.

The mean difference in the onset of the divergence points between the younger and older participants on sentences with verbs that are cognates in both languages was 466 ms, 95% CI [380, 620] ms, and 239 ms, 95% CI [120, 360] ms for non-cognate verbs. These differences are reliable (Fig. 5), indicating that the younger adults predict faster than the older adults when the sentences contain cognate as well as non-cognate verbs. There was a significant difference in prediction between cognate and non-cognate verbs for the younger adults (70 ms; 95% CI [20, 120] ms), but not for the older participants (-157 ms; (95% CI [−340, −0.5]).

Table 5 Bootstrapped divergence points between older and younger participants, for sentences with constraining and non-constraining verbs in each language (2000 iterations). The predictive window is set between 0 and 2050 ms

	Constraining		Non-constraining	
	L1(Norwegian)	L2 (English)	L1 (Norwegian)	L2 (English)
Younger	1184 ms	1188 ms	2153 ms	2160 ms
95% CI	1140–1240 ms	1160–1240 ms	2120–2180 ms	2120–2200 ms
Older	1495 ms	1635 ms	2254 ms	2300 ms
95% CI	1320–1580 ms	1500–1780 ms	2220–2280 ms	2260–2340 ms

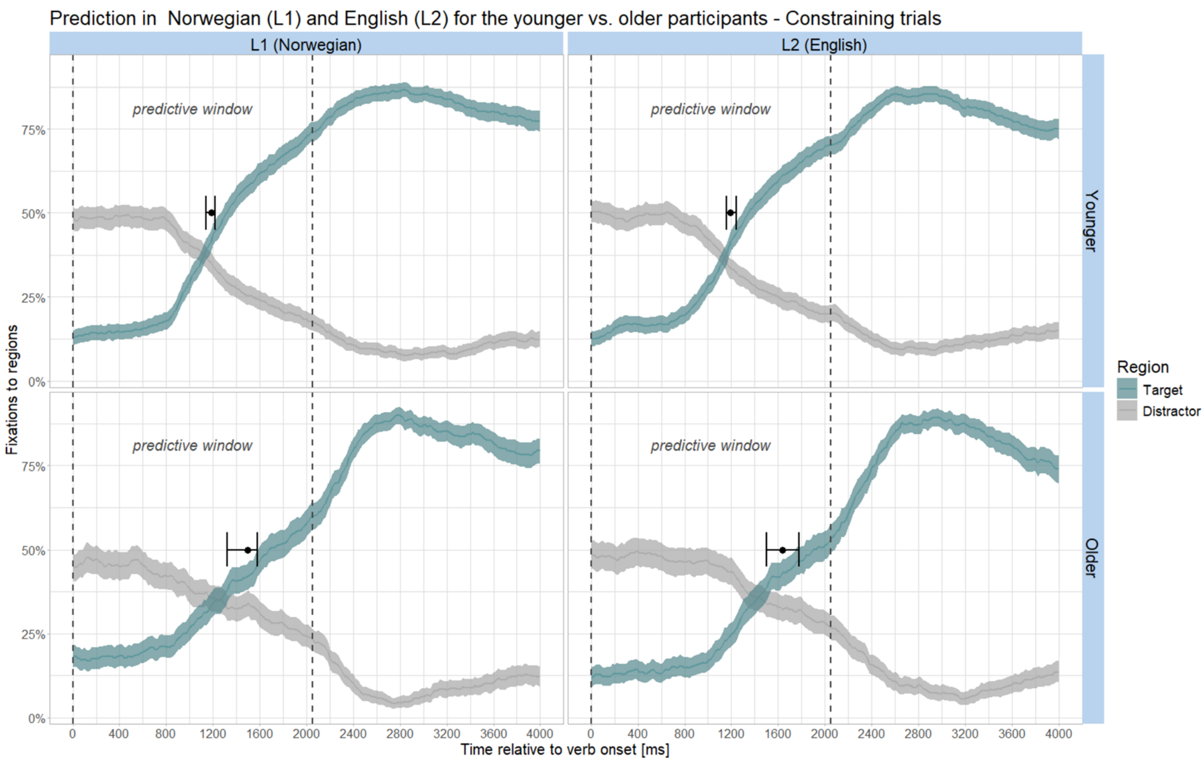

Fig. 2 Fixation curves (proportions of looks to target and distractor) for trials with constraining sentences, in L1 and L2 for younger and older participants. The black dot represents the mean divergence point, and the error bars the 95% confidence intervals (CI) for each group. The shaded area around the fixation curves represents the 95% CI of the non-bootstrapped fixation time-series. The timeline has been recentered to start at the beginning of the predictive window (0 ms), which corresponds to the verb onset. The noun onset is at 1800 ms in all trials, and the predictive window offset is at 2050 ms, allowing for a 250 ms buffer after the noun onset. Proportions of looks to the fixation picture and white space are not plotted

For sentences with cognate nouns in the two languages, the divergence point for the younger participants was at 1130 ms (95% CI between [10810,1180]) after the verb onset and at 1244 ms (95% CI [1200,1280]) after the verb onset for non-cognate nouns, indicating a faster time to detect the target when the nouns were cognates in Norwegian and English. Similarly, the older adults fixated significantly more on the target than on the distractor at 1342 ms (95% CI [1280,1480]) after the onset of the verb for cognate nouns and at 1661 ms (95% CI [1580,1820]) after the verb onset for sentences with non-cognate nouns (Fig. 6).

The mean difference in divergence points between the younger and older participants on sentences with cognate nouns was 212 ms, 95% CI [120, 360] ms, and

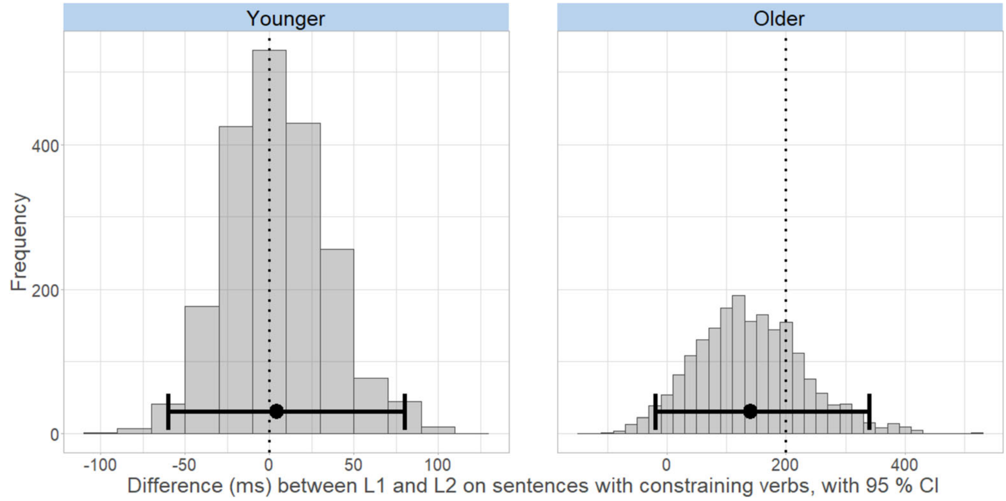

Fig. 3 Bootstrapped divergence point differences between L1 and L2 for younger and older adults. The figure shows the histogram, mean and confidence intervals from the bootstrap, as well as the mean in the original sample (vertical dotted line). In both cases, the confidence intervals include 0, indicating that there is not a reliable difference between the time it takes to detect the target across languages

Table 6 Bootstrapped divergence points between older and younger participants, for cognate and non-cognate verbs and nouns (2000 iterations)

	Verb characteristics		Noun characteristics	
	Cognate	Non-cognate	Cognate	Non-cognate
Younger	1154 ms	1224 ms	1130 ms	1244 ms
95% CI	1120–1200 ms	1180–1260 ms	1081–1180 ms	1200–1280 ms
Older	1619 ms	1462 ms	1342 ms	1661 ms
95% CI	1540–1780 ms	1360–1580 ms	1280–1480 ms	1580–1820 ms

417 ms, 95% CI [320, 580] ms for non-cognate nouns. These differences (Fig. 7) are reliable, indicating a cognate facilitation effect for nouns, in that both younger and older participants fixate more quickly on the target image when the noun is a cognate between English and Norwegian, than when it is a non-cognate. For both the younger and older adults there was a reliable difference in the time it took to detect the cognate noun (younger: 114 ms with the 95% CI [60, 180]; older: 319 ms, with the 95% CI [160, 500]) over the non-cognate noun.

Influence of age, education, language proficiency and working memory on predictive ability

Descriptive results from the cognitive and language tasks were presented in Table 1. To explore if there were any correlations between predictive ability and language or cognitive factors, and between the different background measures, we ran Pearson correlations as exploratory analyses separately for the sentences with constraining verbs in each language. Given the small sample size of our study, we performed a power-determination analysis and an effect-size sensitivity analysis, with the *pwr* package (Champely et al., 2022), post hoc, to estimate the power of our sample and the smallest effect size that can be detected from the reported sample size. Given a sample size of N = 44, an expected medium effect size (r = 0.5) and a significance level at 0.05, the power-determination analysis gives an estimated power of 0.94. The effect-size sensitivity analysis assumes that with a power level of 0.80, the sample size N = 44, and a significance level of 0.05, the smallest effect size that can be

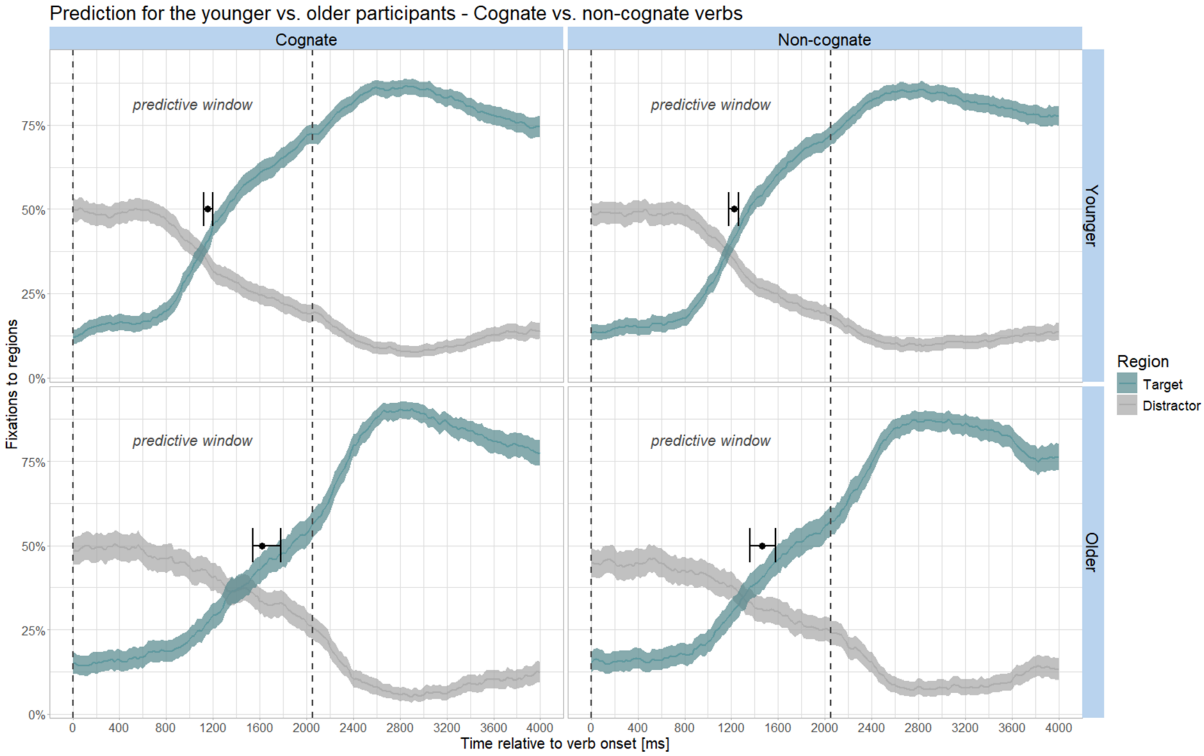

Fig. 4 Fixation curves (proportion of looks to the target and distractor images) for the differences between cognate and non-cognate verbs for the younger and older participants. The black dot represents the mean divergence point, and the error bars the 95% confidence intervals (CI) for each group. The shaded area around the fixation curves represents the 95% CI of the non-bootstrapped fixation time-series. The timeline has been recentered to start at the beginning of the predictive window (0 ms), which corresponds to the verb onset. The noun onset is at 1800 ms in all trials, and the predictive window offset is at 2050 ms, allowing for a 250 ms buffer after the noun onset. Proportions of looks to the fixation picture and white space are not plotted

reliably detected is $r = 0.4$ which corresponds to a medium effect size.

For prediction ability in both languages, there was a moderate, negative correlation between the Flanker effect and predictive looks; $r(42) = -0.42$, $p = 0.004$ for Norwegian and $r(42) = -0.34$, $p = 0.02$ for English. Prediction ability in English was negatively correlated with age (r (42) -0.32, $p = 0.03$). Furthermore, there was a moderate positive correlation between age and the difference in verbal fluency scores between L1 and L2 ($r(42) = 0.35$, $p = 0.01$).

Figure 8 shows the correlations between predictive looks and linguistic and cognitive tests in Norwegian (above) and English (below).

Next, we fitted two multiple linear regressions, with results from the cognitive and linguistic background tests as predictor variables, to see whether linguistic or cognitive factors predicted prediction ability.

The model for predictive ability in Norwegian (predictive looks $\sim$ age + digit span backwards + Flanker effect + category fluency) revealed a significant relationship between the Flanker effect and the proportion of predictive looks ($R^2 = -0.001$, $p = 0.02$), meaning that the participants who had a smaller Flanker effect showed more predictive looks to the target. The full results from the regression analysis can be found in Table 7.

Multicollinearity in the model was explored by calculating the variance inflation factor (VIF) between the different variables in the model. All VIFs were below 2.5, which is often considered a reliable cut-off value (Johnston et al., 2017): age = 1.05, Digit span backwards = 1.04, Flanker effect = 1.12, Norwegian category fluency = 1.08.

In the English model (predictive looks $\sim$ age + digit span backwards + Flanker effect + category

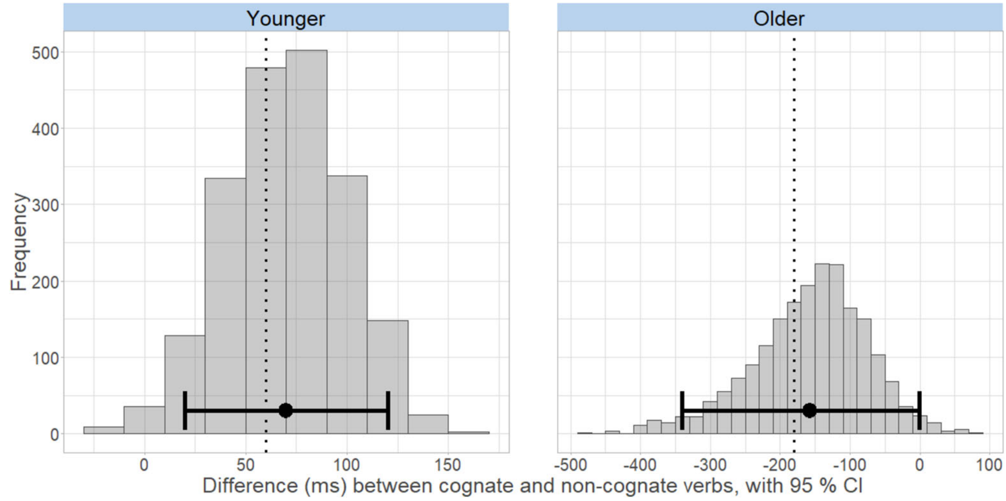

Fig. 5 Bootstrapped divergence point differences between sentences with cognate and non-cognate verbs for younger and older adults in both languages. The figure shows the histogram, mean and confidence intervals from the bootstrap, as well as the mean in the original sample (vertical dotted line). For the younger, the confidence intervals do not include 0, indicating that there is a reliable difference between the timing to detect the target across languages. In the case of the older adults, the confidence intervals range from -340 to -0.5, thus barely inside of 0

fluency + verbal fluency difference + word familiarity + self-rated proficiency), only age was a significant predictor of prediction ability ($R^2 = -0.003$, $p = 0.01$), meaning that the younger participants had more predictive looks to the target image than the older participants. The full results of the regression model can be found in Table 8.

We tested for multicollinearity in the model, by examining the VIF between variables. All variables in the model had a VIF below 2.5: age = 1.48, digit span backwards = 1.41, Flanker effect = 1.11, verbal fluency difference between Norwegian and English = 1.37, English category fluency = 1.62, word familiarity = 1.03, Self-rated proficiency = 1.20.

Discussion

We conducted an eye-tracking experiment with Norwegian-English bilingual younger and older adult participants to examine whether bilingual people predict upcoming information during a sentence listening task. Specifically, we asked whether bilingual participants predicted in both their L1 and their L2, whether there was an age-related difference in predicting abilities, whether individual differences, including language proficiency and working memory abilities, interacted with prediction behavior, and whether cognate status of the verb or of the target noun had an effect on prediction. We address our findings with respect to our research questions and the published literature.

Prediction and bilingualism

Our results revealed that the participants in this study predicted upcoming nouns in sentences with constraining verbs in both Norwegian and English. There was no difference between the prediction behavior of the participants when listening to sentences in their L1 (Norwegian) versus in their L2 (English). This finding is consistent with several studies that showed predictive looks when participants engaged in an eye-tracking experiment in their L2 (Dijkgraaf et al., 2017, 2019). Our findings diverge from the findings of slower or differential prediction patterns in the L2 reported in several studies (e.g., Hopp, 2015; Lew-Williams & Fernald, 2010). Differences across studies could be related to the predictive cues utilized, for example, grammatical or morphological (gender, case) versus semantic cues. The results of our study, which focused on semantic cues, align with those found in previous studies using semantic cues (e.g., Dijkgraaf et al., 2017).

In summary, the answer to our first research question is that our participants, who were highly

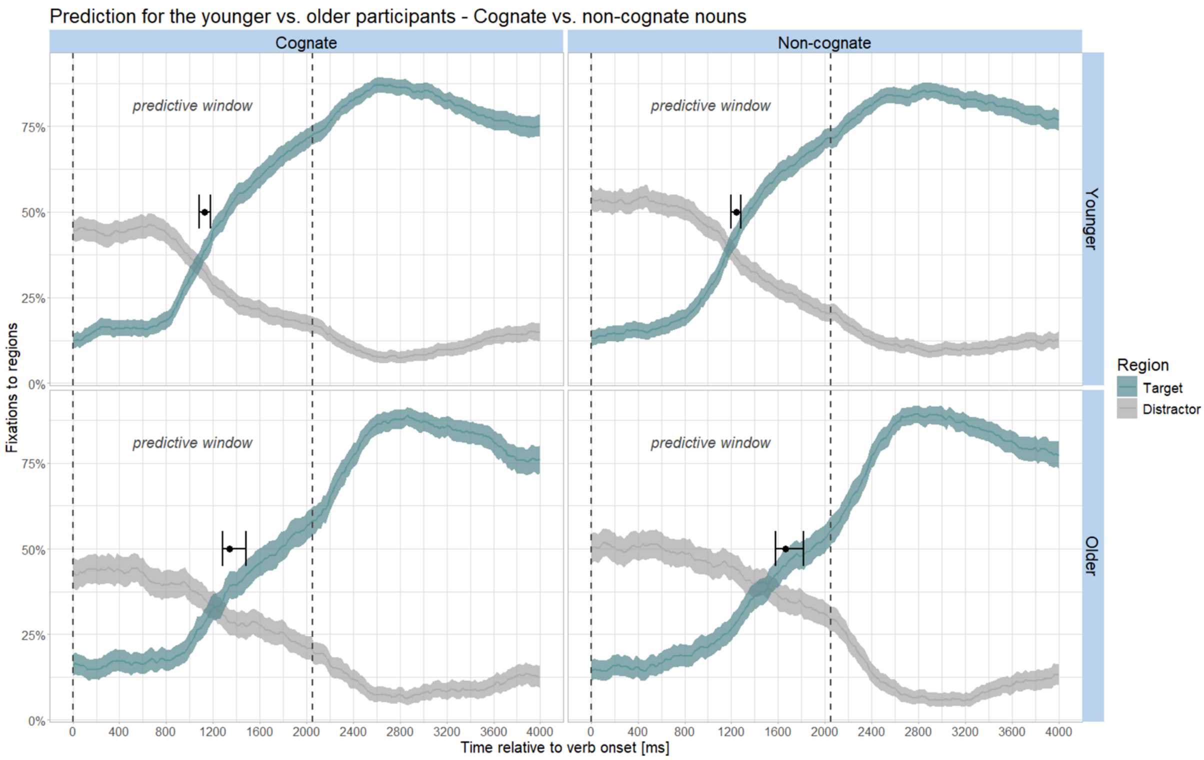

Fig. 6 Fixation curves (proportion of looks to the target and distractor images) for the differences between cognate and non-cognate nouns for the younger and older participants. The black dot represents the mean divergence point, and the error bars the 95% confidence intervals (CI) for each group. The shaded area around the fixation curves represents the 95% CI of the non-bootstrapped fixation time-series. The timeline has been recentered to start at the beginning of the predictive window (0 ms), which corresponds to the verb onset. The noun onset is at 1800 ms in all trials, and the predictive window offset is at 2050 ms, allowing for a 250 ms buffer after the noun onset. Proportions of looks to the fixation picture and white space are not plotted

proficient bilingual people, predicted upcoming nouns following constraining verbs, in both their L1 and L2.

Prediction and aging

Our comparisons of younger and older participant groups revealed a reliable difference between them: while both younger and older participants looked to the target within the predictive window, the older adults did so later than the younger adults. That is, they showed the expected looking patterns but were slower to fixate on the target noun compared to the younger adults. This was true for performance in both the L1 and L2, a finding that has not been reported in previous studies. Our results are consistent with those of Federmeier et al. (2002) who reported less efficient sentence processing by older adults compared to younger adults. Our results are inconsistent with the hypothesis that older adults rely more on top-down processing and thus should show better prediction performance (Federmeier et al., 2010).

Thus, the answer to our second research question is that age affects performance, with older adults looking to the target later than younger adults, but that older adults, like their younger peers, use the constraining information to predict upcoming information in the sentence.

Effects of language proficiency and cognitive abilities

It has been suggested that language proficiency can account for some inconsistent results reported in the literature regarding prediction performance in participants' L2, in accordance with the RAGE hypothesis (Grüter et al., 2014). For example, several studies that

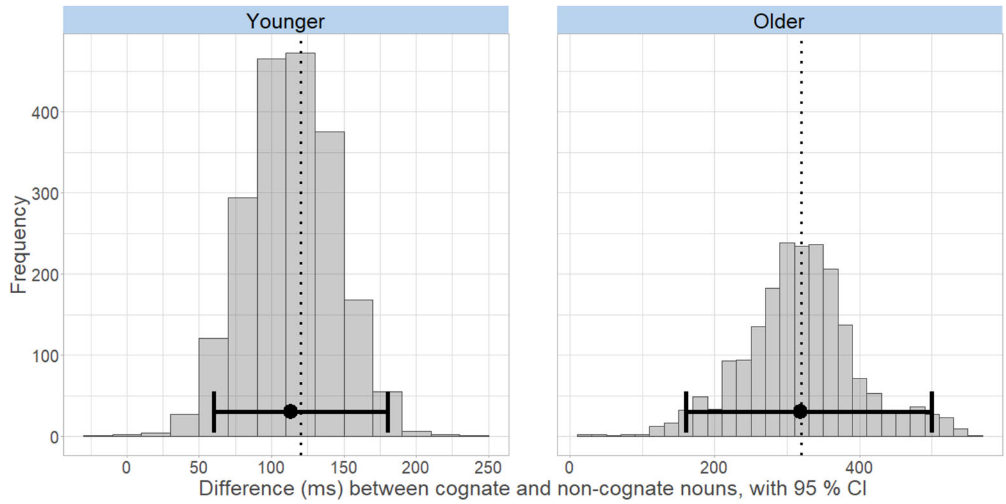

Fig. 7 Bootstrapped divergence point differences between sentences with cognate and non-cognate nouns for younger and older adults. The figure shows the histogram, mean and confidence intervals from the bootstrap, as well as the mean in the original sample (vertical dotted line). In both cases, the confidence intervals do not include 0, indicating that there is a reliable difference between the timing to detect the target across languages

reported divergent prediction performance between L1 and L2 speakers included learners of an L2 as their participants, whereas studies that found no differences between native and non-native prediction abilities involved highly proficient bilingual individuals (e.g., Grüter et al., 2012; Perdomo & Kaan, 2021). The participants in our study were highly proficient in both their languages. Moreover, we confirmed that the participants were familiar with the words we used in the experiment, whereas previous studies did not always report whether the participants knew the critical words.

Lexical knowledge has been suggested to underlie prediction abilities (e.g., Peters et al., 2018). One hypothesis, the “prediction by production” hypothesis, presented by Pickering and Gambi (2018), suggests that an efficient way to predict upcoming information is by constructing the communicative intention of the speaker. To do so, listeners engage their own language production system. Pickering and Gambi (2018) cite evidence that the production system is engaged during prediction in comprehension experiments. In their review paper, Pickering and Gambi (2018) go on to suggest that in certain cases, prediction does not engage production processing, for example, in the case of L2 speakers. This may be due to insufficient proficiency in the languages in which comprehension and prediction are tested, and may lead to less efficient prediction.

Similarly, several studies found a relationship between vocabulary scores and markers of prediction (e.g., Borovsky et al., 2012; Federmeier et al., 2002; Rommers et al., 2015). However, not all studies found such a relationship with vocabulary; for example, Dijkgraaf et al. (2017) did not find a relationship between predictive ability and vocabulary scores of their L2 listeners, but they argued that the vocabulary test they used (LexTALE; Lemhöfer & Broersma, 2012) might not have been sensitive enough to show variation in vocabulary scores. We measured lexical abilities in L2 via a translation task, verbal fluency tasks, and self-rated proficiency. In line with previous studies that did not find a robust relationship between language proficiency measures and prediction behavior (Corps et al., 2023; see Schlenter, 2023), we did not find participants’ L2 proficiency to predict behavior during L2 processing. Recall, however, that all our participants were highly proficient. Results might be different when including participants with more variation in their L2 proficiency.

As discussed in previous studies, it is possible that cognitive abilities, reported to decline in older age, account for the differences in performance between the younger and older participants. Our results demonstrated that although working memory (WM)

Fig. 8 Correlations between cognitive and linguistic scores and predictive looks in Norwegian (L1), top, and English (L2), bottom

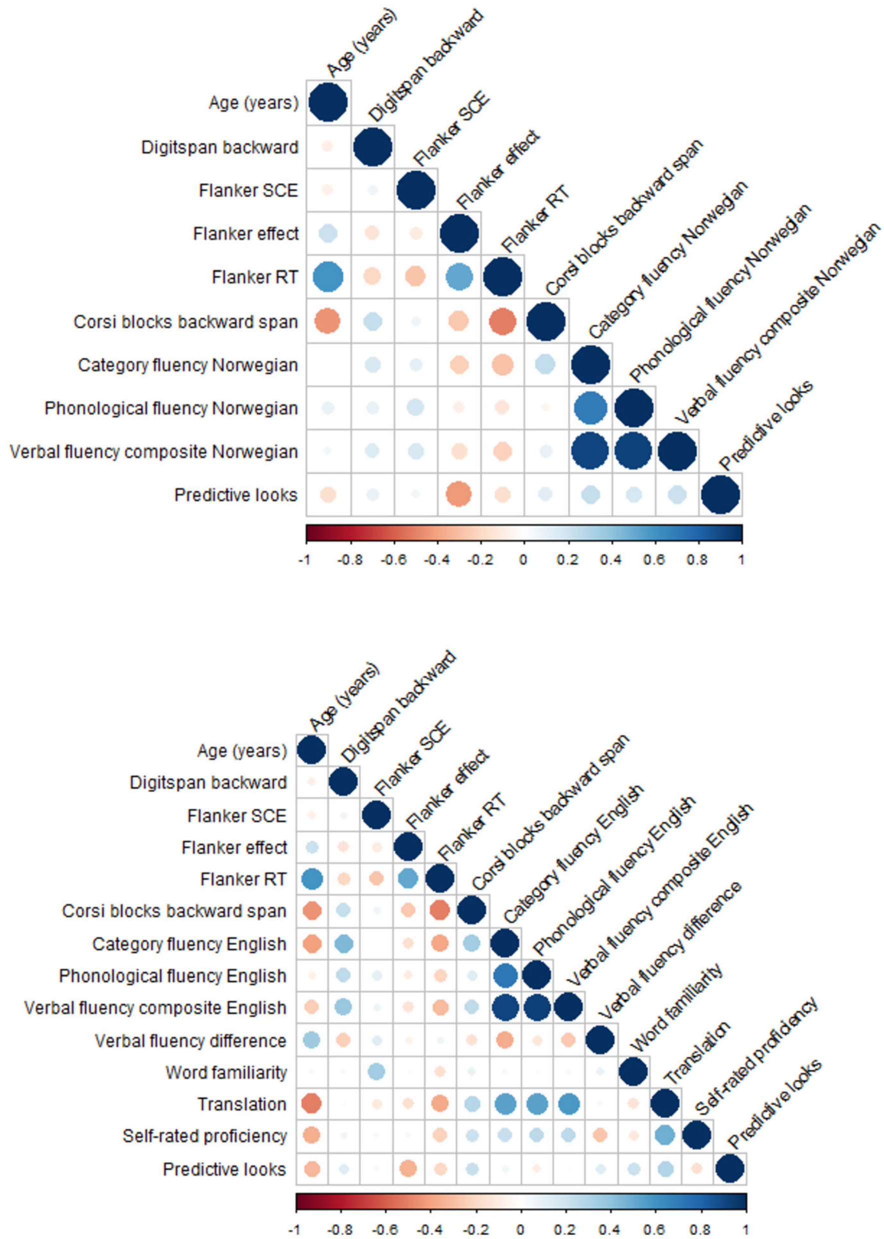

Table 7 Results from the regression model for predictive looks and cognitive and linguistic background measures in L1

	Estimate	sd	t-value	*p*-value
Intercept	0.38	0.14	2.58	0.01*
Age	< -0.001	0.001	−0.55	0.57
Digit span backwards	< -0.001	0.01	0.09	0.92
Flanker effect	−0.001	< 0.001	−2.24	0.02*
Category fluency Norwegian	0.001	0.002	0.95	0.34

*Indicates the significance level ($p < 0.05$)

Table 8 Results from the regression model for predictive looks and cognitive and linguistic background measures in L2

	Estimate	sd	t-value	*p*-value
Intercept	0.63	0.30	2.08	0.04*
Age	−0.003	0.001	−2.52	0.01*
Digit span backwards	0.007	0.01	0.64	0.52
Flanker effect	< -0.001	< 0.001	−1.44	0.15
Verbal fluency difference	0.003	0.003	1.16	0.25
Category fluency english	−0.001	0.002	−0.36	0.71
Word familiarity	< 0.001	< 0.001	0.99	0.32
Self-rated fluency	−0.03	0.02	−1.63	0.11

*Indicates the significance level ($p < 0.05$)

abilities (as measured by the Corsi block test and by a digit span task) correlated with predictive looks to the targets, none of the WM measures we used predicted behavior in the regression analyses. This result contrasts with previous reports (e.g., Li & Qu, 2024), although those studies examined performance in younger, not older, adults. In contrast, we found that our inhibition measure, the Flanker test, predicted the eye-tracking results, though only in participants' L1. These results are consistent with the findings reported in Ito et al. (2018), albeit for a somewhat different cognitive measure. Because we found an age effect in participants' L2 performance and because age correlated with the Flanker effect, it is possible that the relationship was masked in the L2 analysis. Age also correlated with WM, which could have prevented WM to emerge as a predictive variable in the regressions.

Age-related slow speed of processing has been suggested to account for cognitive performance differences between younger and older adults (e.g., Huettig & Janse, 2016; Salthouse, 1991), and could have explained the finding of later divergent points for the older compared to the younger adults in our study. However, our measure of processing speed (reaction time in the Flanker task) did not correlate strongly with predictive looks and thus was not included in our regression analyses. We note that the cognitive measures used here may not capture all potential cognitive changes associated with age and also that there was variability among the older participants' performance. Future studies can further explore the interrelation among measures of inhibition, WM, and speed, and their relation to prediction behavior.

In summary, in response to our third research question about the roles of language proficiency and cognitive abilities in prediction, for highly proficient bilingual people prediction processes may be similar in both their languages. As for cognitive abilities, of the measures used in this study, only the task associated with inhibition skills predicted the eye-tracking results, and only in participants' first language.

Effects of cognate status of the stimuli

Because cognates arguably have stronger lexical representations than non-cognates (e.g., Dijkstra et al., 2010), due to their shared form and meaning, we hypothesized a cognate facilitation effect in prediction in our bilingual participants. For both older and younger participants we found faster prediction in sentences with cognate than non-cognate nouns. The cognate facilitation effect for nouns is in line with the few existing prediction studies involving cognates (e.g., Duyck et al., 2007; Schwartz & Kroll, 2006). The younger adults were also faster to predict in sentences with cognate verbs than with non-cognate verbs. However, this effect is smaller than the cognate effect for nouns. We found no cognate facilitation effect for verbs for the older participants. These results align with the findings by Van Assche et al. (2013), who found weaker cognate facilitation effects in verbs than in nouns, and argued that weaker cognate effects may be attributed to a more demanding processing of verbs than nouns due to semantic, syntactic and morphological differences between the two word classes. However, our results are not completely compatible with those of Van Assche et al. (2013), since in their study the verb was the target to be predicted, and the eye tracking tasks are not quite comparable.

Indeed, another reason for the different results for verbs and nouns can be found in the different roles of the two word classes in our study: the verbs functioned as cues to prediction, whereas the nouns were the words to be predicted. Thus, the verb was actually heard by the participants before prediction took place, already facilitating access, so here the cognate status might play a lesser role. The noun, on the other hand, was not heard within the predictive window, and a stronger lexical representation would have a more important role in contributing to prediction.

In summary, in response to our fourth research question, form and meaning similarities between the to-be-predicted nouns facilitated the prediction process, but for the cognate verbs functioning as a cue, the facilitation effect was smaller and limited to the younger age group.

Limitations and future directions

Our study is limited by the modest sample size, especially among the older adults. The low degree of prediction found for the older participants should be interpreted with caution, and further studies are needed to explore the effects of aging on linguistically mediated prediction. Great variation has been reported for older adults and for bilingual adults in cognitive test performance (e.g., Sliwinski & Buschke, 2004), which may limit the generalizability of our current findings. Future studies with additional participants are needed to corroborate our results. In addition, investigating prediction performance across varying language combinations among bilingual people will further our understanding of prediction behavior. For example, varying predictive cues (grammar-based cue, semantic cue) may weigh differently across languages and similarities and differences in such cues between the two languages of bilingual people can affect predictive behavior in their L2.

Analyzing eye tracking data with a divergence point analysis with bootstrapping can give reliable information about when people look at different images on a screen. While building our bootstrapping statistic function, three different base tests were considered, and we chose a t-test to compare sample means over binomial tests and linear regressions with weighted logits. Future studies could provide further insight into the type of test that is best suited for eye-tracking data when building a bootstrapping model.

Furthermore, we found a cognate effect for nouns, but for verbs, only for the younger participants. While these results are somewhat in line with the few earlier studies documenting cognate facilitation in prediction among bilingual people (e.g., Duyck et al., 2007; Van Assche et al., 2013), this aspect of prediction merits further investigation. For instance, looking at cognateness as a continuous variable to identify differences between different degrees of cognateness may be a way forward.

Finally, not all of our participants completed the online translation task (eight younger and three older adults did not), which may be considered our more objective measure of language proficiency. Nevertheless, our participant sample was characterized by high L2 proficiency. As discussed above, proficiency level variation can potentially account for inconsistencies in the literature on prediction behavior in bilingual people. Additional studies could contribute to the unpacking of language proficiency effects on bilingual language prediction processes.

Conclusion

This study utilized eye-tracking experiments to examine semantically based prediction behavior of younger and older highly proficient Norwegian-English bilingual people in both their languages. The results highlight equal prediction abilities in the first and second languages for both younger and older participants, suggesting that bilingual people do predict upcoming semantic information when listening to sentences in both their L1 and L2. Older adults were slower than younger adults, as they looked to the target images towards the end of the predictive window. Language proficiency did not explain performance, and cognate status did so only partly, though we note that these null effects may stem from limited variability on the proficiency measures in the current sample and/or low power. Future studies examining other language pairs and differing proficiency levels, as well as different word classes and other constraining linguistic information, will complement the current findings.

Acknowledgements Sarah Cameron, Anne Marte Haug Olstad, Cristiano Cardone, Jessica Pedersen Belisle Hansen, Michela Iacorossi, Henriette Johansen, and Ekaterina Kuzmina for their contributions, as well as the participants who took part in the study.

Author contributions Authors H.G.S., M.G., M.L. and M.N. made substantial contributions to the conception or design of the work, and authors A.T., I.S.R and J.M. substantially contributed to the acquisition of the data. Author T.N. contributed substantially to the analyses together with I.S.R., The results were interpreted by H.G.S., I.S.R., M.G., M.L. and M.N. and T.N. T.N. created the customized 'boot' function and prepared Figs. 2, 3, 4, 5, 6 and 7 in the manuscript. I.S.R. prepared Figs. 1 and 8. The first draft of the manuscript was prepared by H.G.S. and M.G., and authors A.T., I.S.R. and J.M. contributed with writing specific sections related to methods, materials, analysis and results. All authors edited and reviewed the manuscript.

Funding Open access funding provided by OsloMet - Oslo Metropolitan University. The study was supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223265.

Data availability Preprocessed eye tracking data, as well as data from linguistic and cognitive background tests are available on OSF (https://doi.org/10.17605/OSF.IO/VEMY4). The study was not preregistered.

Declarations

Conflict of interest The authors declare no competing interests.

References

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2024). rmarkdown: Dynamic Documents for R. R package version 2.28. https://github.com/rstudio/rmarkdown. Accessed 29 Aug 2024

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

Audacity Team. (2018): Audacity(R): Free Audio Editor and Recorder [Computer Software]. Version 2.2.2. Available online at: https://audacityteam.org/. 7 Nov 2024.

Barr, D. (2008). Analyzing 'Visual World' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barrett, T,. Dowle, M,. Srinivasan, A., Gorecki, J., Chirico, M., & Hocking, T. (2024). data.table: Extension of 'data.frame'. R package version 1.15.4, https://CRAN.R-project.org/package=data.table. Accessed 29 Aug 2024.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bengtsson, H. (2023). R.utils: Various Programming Utilities. R package version 2.12.3, https://CRAN.R-project.org/package=R.utils. Accessed 29 Aug 2024.

Blumenfeld, H. K., Bobb, S. C., & Marian, V. (2016). The role of language proficiency, cognate status and word frequency in the assessment of Spanish–English bilinguals' verbal fluency. *International Journal of Speech-Language Pathology, 18*(2), 190–201.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Canty, A., & Ripley, B. (2024). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-29.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2022). pwr. R package version 1.3-0, https://cran.r-project.org/web/packages/pwr/ Accessed 29 Aug 2024.

Contemori, C., & Dussias, P. E. (2019). Prediction at the discourse level in Spanish-English bilinguals: An eye-tracking study. *Frontiers in Psychology, 10*, 956.

Corps, R. E., Liao, M., & Pickering, M. J. (2023). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition, 26*, 231–243. https://doi.org/10.1017/S1366728922000499

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283.

Csárdi, G. & FitzJohn, R. (2023). progress: Terminal Progress Bars. R package version 1.2.3. https://CRAN.R-project.org/package=progress. Accessed 29 Aug 2024.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge University Press.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition, 20*(5), 917–930.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience, 34*(7), 881–900.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language, 62*(3), 284–301. https://doi.org/10.1016/j.jml.2009.12.003

Donnelly, S., & Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. *Journal of Memory and Language, 94*, 28–42. https://doi.org/10.1016/j.jml.2016.10.005

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115*, 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133–146.

Fox, J. (2019). An R companion to applied regression, third edition. Sage, Thousand Oaks, California. https://socialsciences.mcmaster.ca/jfox/Books/Companion/. Accessed 29 Aug 2024.

Garrido Rodriguez, G., Norcliffe, E., Brown, P., Huettig, F., & Levinson, S. C. (2023). Anticipatory processing in a verb-initial Mayan language: Eye-tracking evidence during sentence comprehension in Tseltal. *Cognitive Science, 47*, e13292. https://doi.org/10.1111/cogs.13219

Giner-Sorola, R., Montoua, A. K., Reifman, A., Carpenter, T., Lewis, N. A., Jr., Aberson, C. L., Bostyn, D., Conrique, B. G., Ng, B. W., Schoeman, A. M., & Soderberg, C. (2024). Power to detect what? Considerations for planning and evaluating sample size. *Personality and Social Psychology Review, 28*(3), 276–301. https://doi.org/10.1177/10888683241228328

Grundy, J. G. (2022). The Specificity and reliability of conflict adaptation: A mouse-tracking study. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.770509

Grüter, T., Rohde, H., & Schafer, A. (2014). The role of discourse-level expectations in non-native speakers'referential choices. In: Proceedings of the annual Boston university conference on LANGUAGE DEVELOPMENT.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191–215.

Harrell, F. E., Jr. (2024). Hmisc: Harrell Miscellaneous. R package version 5.1-3, https://CRAN.R-project.org/package=Hmisc. Accessed 29 Aug 2024.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53*(3), 277–306.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition, 21*(2), 251–264. https://doi.org/10.1017/S1366728917000050

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods, 55*, 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language, 44*(1), 2–31.

Johnston, R., Jones, K., & Manley, D. (2017). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality and Quantity, 52*, 1957–1946.

Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition, 43*(1), 133–154.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*, 325–343.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63*(4), 447–464.

Li, X., & Qu, Q. (2024). Verbal working memory capacity modulates semantic and phonological prediction in spoken comprehension. *Psychonomic Bulletin & Review, 31*, 249–258.

Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B. H. (2015). Norwegian Words: A lexical database for clinicians and researchers. *Clinical Linguistics & Phonetics, 29*(4), 276–290.

Lo, C. H., Hermes, J., Kartushina, N., Mayor, J., & Mani, N. (2023). e-Babylab: An open-source browser-based tool for unmoderated online developmental studies. *Behavioral Research Methods*. https://doi.org/10.3758/s13428-023-02200-7

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847. https://doi.org/10.1037/a0029284

Maquate, K., & Knoeferle, P. (2021). Referential vs. non-referential world-language relations: How do they modulate language comprehension in 4 to 5-year-olds, younger, and older adults? *Frontiers in Psychology, 11*, 3248.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research JSLHR, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martínez-García, M. T. (2019). Using eye-movements to track bilingual activation. *Languages, 4*, 59. https://doi.org/10.3390/languages4030059

McConnell, K., & Blumenthal-Dramé, A. J. (2021). Usage-based individual differences in the probabilistic processing of multi-word sequences. *Frontiers in Communication*. https://doi.org/10.3389/fcomm.2021.703351

Pelak, V. S. (2010). Ocular motility of aging and dementia. *Current Neurology and Neuroscience Reports, 10*, 440–447. https://doi.org/10.1007/s11910-010-0137-z

Perdomo, M., & Kaan, E. (2021). Prosodic cues in second-language speech processing: A visual world eye-tracking study. *Second Language Research, 37*(2), 349–375.

Peters, R., Grüter, T., & Borovsky, A. (2018). Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied Psycholinguistics, 39*(6), 1439–1469. https://doi.org/10.1017/S0142716418000383

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*, 1002–1044. https://doi.org/10.1037/bul0000158

Posit team (2024). RStudio: integrated development environment for R. Posit Software, PBC, Boston, MA. http://www.posit.co/. Accessed 29 Aug 2024.

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 29 Aug 2024.

Ramscar, M., & Baayen, H. (2015). Learning and cognitive maturation. *The Encyclopedia of Adulthood and Aging*. https://doi.org/10.1002/9781118521373.wbeaa317

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6*(1), 5–42. https://doi.org/10.1111/tops.12078

Revelle, W. (2024). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.4.3, https://CRAN.R-project.org/package=psych. Accessed 29 Aug 2024.

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics, 77*(3), 720–730.

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2*(3), 179–183. https://doi.org/10.1111/j.1467-9280.1991.tb00127.x

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language, 162*, 42–45.

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition, 26*, 253–267. https://doi.org/10.1017/S1366728922000736

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55* (2), 197–212. https://doi.org/10.1016/j.jml.2006.03.004

Scott, G. G., Keitel, A., Becirspahic, M., et al. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods, 51*, 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

Sliwinski, M., & Buschke, H. (2004). Modeling intraindividual cognitive change in aging adults: Results from the Einstein aging studies. *Aging, Neuropsychology, and Cognition, 11*(2–3), 196–211. https://doi.org/10.1080/13825580490511080

Spreen, O. & Benton, A. L. (1969). Neurosensory center comprehensive examination for aphasia: Manual of instructions (NECCEA). Victoria.

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analysis of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition, 24*, 833–841. https://doi.org/10.1017/S1366728920000607

Strangmann, I., Antolovic, K., Hansen, P., & Simonsen, H. G. (2023). The cognate continuum: Approaches to empirically establishing form overlap. *The Mental Lexicon, 18*(1), 151–175.

Swinburn, K., Porter, G., Howard, D., Høeg, N., Norvik, M., Røste, I., & Simonsen, H. G. (2021). *CAT-N: Comprehensive aphasia test. Norwegian version*. Novus Forlag.

Van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts: The effect of cognate status and verb tense. *Studies in Second Language Acquisition, 35*(2), 237–259.

Vanek, N., Matić Škorić, A., Košutar, S., Matějka, Š, & Stone, K. (2024). Looks at what isn't there: eye movements on a blank screen when processing negation in a first and a second language. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2024.1457038

Veríssimo, J., & Clahsen, H. (2014). Variables and similarity in linguistic generalization: Evidence from inflectional classes in Portuguese. *Journal of Memory and Language, 76*, 61–79. https://doi.org/10.1016/j.jml.2014.06.001

Waris, O., Soveri, A., & Laine, M. (2015). Transfer after Working Memory Updating Training. *PLoS ONE, 10*(9), e0138734. https://doi.org/10.1371/journal.pone.0138734

Wei, T. & Simko, V. (2021). R package 'corrplot': Visualization of a correlation matrix (Version 0.92). https://github.com/taiyun/corrplot. Accessed 29 Aug 2024.

Wickham H., Pedersen T., & Seidel D. (2023). scales: Scale functions for visualization. R package version 1.3.0, https://CRAN.R-project.org/package=scales. Accessed 29 Aug 2024.

Wickham, H., Girlich, M., Fairbanks, M., & Dickerson, R. (2023). dtplyr: Data table back-end for 'dplyr'. R package version 1.3.1, https://CRAN.R-project.org/package=dtplyr. Accessed 29 Aug 2024.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes, 28*(3), 272–302. https://doi.org/10.1080/01690965.2011.622905